Predicting Chronic Kidney Disease Progression from Stage III to Stage V using Language Models

Zainab Awan
NIHR Newcastle PSRC

Authors: Zainab Awan, Rafael Henkin, Nick Reynolds†, Michael R. Barnes

Background: Chronic Kidney Disease (CKD) affects 5 – 10 % of the global population. Progression from Stage III (eGFR 30–59 mL/min/1.73 m²) to Stage V (< 15 mL/min/1.73 m²) markedly increases morbidity and mortality, yet remains difficult to predict with existing clinical tools. Early, reliable prediction could enable timely interventions that slow decline and improve outcomes.

Objectives: Can transformer-based language models, when fed routinely collected primary care data, accurately predict which Stage III CKD patients will progress to Stage V? How do full finetuning and parameterefficient LowRank Adaptation (LoRA) compare for this task?

Methods:
Cohort: CPRD GOLD records (2010 – 2020) of adults (16 y) with 2 longterm conditions. CKD≥≥ stages identified via READ v2/ICD10 codes.

Sample: After exclusions and 1:1 age-matched propensity scoring, 4 606 Stage III (non-progressors) and 4 606 Stage V (progressors) patients.

Inputs: Ethnicity, longterm conditions, and “continuous prescriptions” (3 prescriptions with≥ 84 days between fills) prior to Stage III diagnosis, tokenised into sequences.≤

Models: Random Forest, XGBoost, and transformer encoders (BioBERT, ClinicalBERT, SciBERT, GatorTron, BlueBERT, etc.) trained with (i) full finetuning (FFT, 3 epochs) and (ii) LoRA (r = 16, 5 epochs) using AdamW (LR 2 × 10^{-5}, weight decay, gradient clipping 1.0).

Evaluation: 5-fold stratified cross-validation; metrics: Accuracy, Precision, Recall, F1, AUC; best model selected by validation loss.

Results:
BioBERT (FFT) achieved the highest performance: AUC 0.7787, Precision 0.7261, Accuracy 0.7045, outperforming all other transformers and traditional models.

LoRA variants were computationally lighter but consistently lagged behind full finetuning across all metrics.

Conclusion: Domain-specific full finetuning of BioBERT offers the most accurate prediction of CKD progression from Stage III to V within primary-care data, whereas LoRA sacrifices performance for efficiency. Future work will explore temporal models, model interpretability (e.g., SHAP), and validation in external cohorts.

Predicting Chronic Kidney Disease Progression from Stage III to Stage V using Language Models

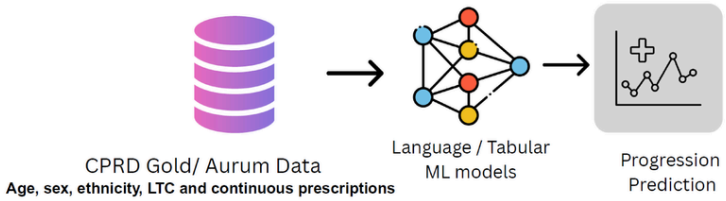
Zainab Awan[1], Rafael Henkin[1], Nick Reynolds[2], Michael R. Barnes[1,2]

1 Center for Translational Bioinformatics, William Harvey Research Institute, Queen Mary University of London, UK
2 Translational and Clinical Research Institute and NIHR Biomedical Research Centre, Newcastle University, Newcastle upon Tyne, UK

NIHR SafetyNet Symposium 2025
psrc-network.nihr.ac.uk

Patient Lay Summary

Chronic Kidney Disease (CKD) happens when the kidneys slowly lose their ability to clean the blood. We looked at people whose kidney function was already reduced (Stage III) to see who might get much worse (Stage V). We used information from GP records, such as other illnesses, medicines, and ethnicity and tested advanced computer programs called "language models" (similar to ChatGPT) to predict who would get worse. The best model correctly identified about 70 out of every 100 patients. This could help doctors spot high-risk patients earlier and provide treatments to slow down the disease.



Results

- Best model: BioBERT (FFT) - AUC 0.7787, Precision 0.7261, Accuracy 0.7045, F1 0.6890, Recall 0.6622.
- LoRA variants were computationally efficient but lagged in all metrics.
- Tabular models (RF, XGBoost) performed competitively (AUC ~0.767) but below best transformer models.
- Recall remains a limitation - potential improvements include Mixture of Experts architectures.

Background

Chronic Kidney Disease (CKD) affects 5–10% of the global population. Progression from Stage III (eGFR 30–59) to Stage V (<15) increases morbidity and mortality yet remains hard to predict. Current tools rely on biochemical and demographic data, often with missing values and limited use of unstructured data.

Aims

Can transformer-based language models using primary care data predict which Stage III CKD patients will progress to Stage V? How do Full Fine-Tuning (FFT) and Low-Rank Adaptation (LoRA) compare?

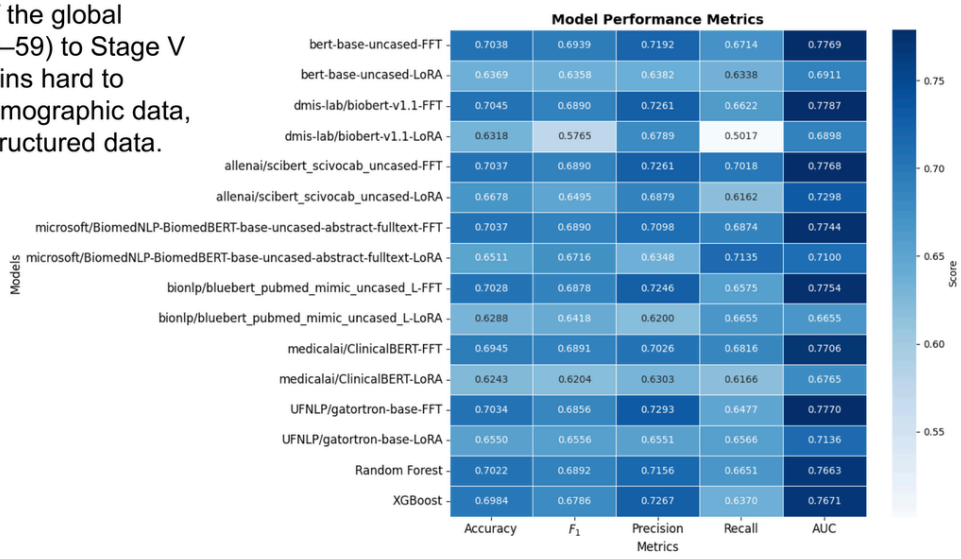


Figure 1: Heatmap of performances across various metrics.

Methods

Data: CPRD GOLD records (2010–2020), adults (≥16y) with ≥2 LTCs. CKD stages identified via READ v2/ICD10. After exclusions & 1:1 age-matched propensity scoring: 4,596 progressors (Stage V) & 4,596 non-progressors (Stage III).
Inputs: Ethnicity, LTCs, and continuous prescriptions prior to Stage III diagnosis.
Models: Random Forest, XGBoost, and transformer encoders (BioBERT, ClinicalBERT, SciBERT, GatorTron, BlueBERT).
Training: FFT (3 epochs) vs LoRA (5 epochs, r=16), AdamW optimiser, LR=2e-5.
Evaluation: 5-fold CV; metrics: Accuracy, Precision, Recall, F1, AUC.

Conclusion

- Domain-specific FFT BioBERT outperformed all other models, demonstrating potential for accurate CKD progression prediction from routinely collected primary care data.
- LoRA offers efficiency but at a cost to predictive performance.
- Future work: temporal models, interpretability (SHAP), validation in external cohorts, and integration into clinical workflows.

z.awan@qmul.ac.uk

NIHR | Newcastle Patient Safety Research Collaboration

This study is funded by the National Institute for Health and Care Research (NIHR) Newcastle Patient Safety Research Collaboration and the NIHR-AIM AI-Multiply Project. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. This research utilised Queen Mary's Apocrita HPC facility, supported by QMUL Research-IT.